DOCUMENT RESUME

ED 165 717 IR 006 281

AUTHOR Uffelman, Robert L.
TITLE Computer Assisted Testing at the Education Resource Center.
PUB DATE Mar 76
NOTE 12p.; Paper presented at the Annual Meeting of the Association for the Development of Computer Based Instructional Systems (Dallas, Texas, March 1-4, 1978); For related document, see ED 160 072 ; Best copy available

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Computer Assisted Instruction; Computer Managed Instruction; Higher Education; *Instructional Design Item Analysis; Systems Development; *Testing
IDENTIFIERS *Computer Assisted Testing

ABSTRACT

The development of the Computer Assisted Testing (CAT) System at the University of Delaware is described. The introduction presents the background leading up to interactive terminal testing in 1973. Documentation for the system includes CAT System programs, format of questions for constructing test item pools, format for entering class lists, testing conditions for terminal tests, sample hard copy test, sample student record, sample anonymous student suggestions, and sample item analysis report. Results of using the CAT System are mentioned briefly. (VT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165717

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COMPUTER ASSISTED TESTING AT THE EDUCATION RESOURCE CENTER

by

Robert L. Uffelman

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

John J. Hirschbuhl

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR006281

COMPUTER ASSISTED TESTING AT THE EDUCATION RESOURCE CENTER

PAPER PRESENTED TO
ADCIS 1978 ANNUAL CONFERENCE IN DALLAS, TEXAS

By
Robert L. Uffelman
University of Delaware, Newark, Delaware 19711

INTRODUCTION

Alternative strategies were investigated by faculty to provide more efficient instruction to increasing numbers of students who were expected to learn more and to retain the knowledge longer. Traditional quiz sessions and hourly examinations did not provide for individualizing student progress nor did they provide for additional study and re-testing that would lead to more learning by more students.

Faculty attempted administration of additional group tests during make-up seminars conducted by teaching assistants. This procedure developed scheduling conflicts for both students and instructors. It also raised the issue of comparability of test items used in different sessions. Security of the item pool disintegrated when corrected tests were returned to subjects for discussion or for other feedback. Generating several printed forms of each test from larger item pools appeared as a promising solution.

Procedures for test generation, scoring and statistical treatment were implemented on the computer. Test items were selected randomly, printed as individual tests and items were replaced in the pool for possible inclusion on other forms of the same test. Thus a subject may or may not see the same item at a subsequent testing session. Items were analyzed statistically and compared for level of difficulty. Face validity was approximated by identifying each item with a concept number keyed to a passage in the textbook, supplemental readings, lectures, or laboratory exercises. Proctors could administer and score different tests while the teaching assistants and faculty conducted elective review sessions. This procedure provided subjects with more rapid feedback but created a major clerical burden when used with large classes. It is still used in two courses (Marler, 1975).

Interactive terminal testing was introduced in 1973 to reduce clerical errors and to improve reporting procedures. Utility programs and subroutines were written, debugged and

revised as faculty investigated several questions about testing behavior and item construction.

Table 1

UTILIZATION OF COMPUTER ASSISTED TESTING SYSTEM

Year	Number of Hard Copy Tests Generated	Number of Interactive Terminal Tests	Total
1972-73	56	428	484
1973-74	4,891	4,428	9,319
1974-75	6,345	14,107	20,452
1975-76	6,998	11,039	18,037
1976-77	4,435	14,926	19,361

More terminals were added but hard copy tests were retained for backup in case of equipment failure or overload as course deadlines were established (Martuza, 1975). Although progress was limited to a single semester by University regulations, greater flexibility was created for many subjects. During 1977-78 test administration routines were combined into a single computer program and utility programs were grouped in a second package. Prompts and cues in the programs permit the testing staff and faculty to perform tasks formerly reserved for the programmmers and system manager without risking loss of files or records. Programmers devote more time to developing experimental routines to handle additional research question (see Figure 1).

PROCEDURES

At the beginning of each semester, faculty using the CAT System provide test questions with correct responses to the testing staff who create a working file. Draft copies are sent to the faculty for editorial changes or corrections.

Figure 1

CAT SYSTEM PROGRAMS

1. Testing Program --

   CATTEST/SUPERVISOR

2. Utility Programs --

SEQ -- SEQUENCER	TSC -- TESTSCORES
UPD -- UPDATEDIR	SUG -- SUGGESTIONBOX
PAC -- PACKER	NEWS --
QDM -- QUESTIONDUMPER (CARDS)	MCB -- MCTESTRECORDSBATCH
TGN -- TESTGENER	TSB -- TESTSCORESBATCH
ENT -- ENTERNAMES	MCT -- MCTEST TOTALANALYSIS
ROS -- CLASSROSTER	CRT -- CRTEST TOTALANALYSIS
MCR -- MCTESTSTUDENTRECORDS	HTR -- HISTORYTRACE
CRR -- CRTESTSTUDENTRECORDS	HRE -- HISTORYRELINK

Figure 2

FORMAT OF QUESTIONS

FOR CONSTRUCTING TEST ITEM POOLS

```
1670000   '035F
1672000   WHICH OF THE FOLLOWING IS/ARE STUDENT PRODUCTS OCCURRING IN A
1674000   NATURAL CONDITION
1676000   A.  EVIDENCE OF DEFACEMENT OF SCHOOL PROPERTY
1678000   B.  AMOUNT OF LITTER ON SCHOOL GROUNDS
1680000   C.  AMOUNT OF USAGE WEAR ON CERTAIN GYM EQUIPMENT AS AN INDEX OF
1682000       PREFERENCE FOR CERTAIN GAMES
1684000   D.  A + B
1686000   E.  B + C
1688000   F.  A + B + C
1690000   G.  NONE OF THE ABOVE
1714000   &10013 ****** END OF ED607 PO *******
```

Seq #167000 Item Number and Correct Response

167200 - 1690000 Item

171400 nnn13 Number of questions to be administered for tests in this unit

The final editions of all items pools are transferred to disk pack for use by subjects. Class lists and identification numbers are entered with testing options assigned to each student.

Figure 3

FORMAT FOR ENTERING CLASS LISTS

```
ENTER THE PROPER ITEMS LEFT JUSTIFIED IN THE FIELDS BORDERED BY '*'.
TO SEE THE GUIDE AGAIN, TYPE $GUIDE.
WHEN YOU HAVE FINISHED, TYPE ?END.

 LAST  NAME   FIRST  SSN COURSE     COURSE     COURSE     COURSE
*----------**-------**--**----*    *----*     *----*     *----*

GUEST        ONE       1111PY301Q
GUEST        TWO       2222ED660Q
```

Four testing conditions for terminal tests have been developed to specify how the supervising program is to display questions, record student responses and display the test session summary after each trial. Four selection options specify how test items are selected from the pool. These options are summarized in Figure 4.

Proctors generate hardcopy tests on either the line printer of the Remote Job Entry (RJE) station in the Education Resource Center or on the Xerox printer at the Computing Center. Each test is labeled by course, unit and sequential number (see Figure 5). The answer strips are separated and tests are filed for future use in a standard cabinet within the testing room. Students write their responses in a separate answer sheet and return their tests and answer sheets to the proctor for correcting (see Figure 6).

Student Record files are updated each day to include any hardcopy tests and are printed at the RJE as insurance that computer problems do not result in lost files. These are kept in a chronological file for future reference by faculty or staff. A sample Student Record file is shown in Figure 7 with Student Comments for faculty members to use in revising items or evaluating the System performance.

Figure 4

TESTING CONDITIONS FOR TERMINAL TESTS

Reinforcement Conditions	OPTIONS			
	Complete		70%	
	WR	R	WR	R
"NORMAL": informs the subject if his answer was correct or incorrect. It forces subjects to respond until the correct answer is obtained. Test score is calculated as number correct on first attempt and is displayed at end of trial.	N	R	A	E
"EXAM": the test is given straight through with only one response to each question. Subjects may not go back and change answers. Scores are displayed at the end of the trial.	O	S	B	F
"POST-TEST REVIEW": the subject may go back to review and change answers after all questions have been given. Score is based on final choice and is displayed at end of trial. The subject may elect to review items to learn correct response after score is displayed.	P	T	C	G
"QUICK REVIEW": the subject may go back at any point in the test to review questions and to change answers. Score is based on final choice and is displayed at the end of the trial. Subject may elect to review items to learn the correct response after score is displayed.	Q	U	D	H

The R, or replacement, option permits the same question to be given in another trial. The WR, or without replacement, option assures that none of the questions will be repeated in any subsequent trials by an individual student. The Complete Selection option generates items for subjects on each trial. The 70% option generates items, but terminates subjects at any time when they miss too many items to permit scoring 70% on a test.

Figure 5

SAMPLE HARD COPY TEST

ED461/ED660 PRETEST FALL 1977

NAME_____ SECTION_____ TIME/TRIAL_____

ED660 NO.= 10
UNIT= 1
ANSWERS

ED660 UNIT= 1 TEST NO.= 10

1

1. THE STATEMENT "JOHNNY IS A BETTER STUDENT THAN MARK" IS ESSENTIALLY -----.
   A. CRITERION-REFERENCED
   B. NORM-REFERENCED

(14,B)

2

2. IF THE HEIGHTS OF 10 INDIVIDUALS WERE MEASURED TO THE NEAREST INCH AND EACH PERSON'S HEIGHT SCORE WERE REDUCED BY 5 INCHES, THE RESULTING SCORES WOULD BEST BE CLASSIFIED AS ----- LEVEL MEASUREMENTS.
   A. NOMINAL
   B. ORDINAL
   C. INTERVAL
   D. RATIO

(8,C)

Figure 6

SAMPLE ANSWER SHEET

NAME:

Course:	#:
Date:	
Score:	Attempt:

-	1)	(26	-
-	2)	(27	-
-	3)	(28	-
-	4)	(29	-
-	5)	(30	-
-	6)	(31	-
-	7)	(32	-
-	8)	(33	-
-	9)	(34	-
-	10)	(35	-
-	11)	(36	-
-	12)	(37	-
-	13)	(38	-
-	14)	(39	-
-	15)	(40	-
-	16)	(41	-
-	17)	(42	-

Figure 7

SAMPLE STUDENT RECORD

NAME: GUEST, ONE IDENTIFICATION: 1111
COURSE: FN211 TEST: 01 TRIAL: 1 UNIT NUMBER: 00
TEST DATE: 09/28/77 TIME ON: 16:41:25 TIME OFF: 12:05:13
RECOVEPED

QUESTION NUMBER	ITEM NUMBER	CORRECT RESPONSE	RESPONSES	CORRECT	TRIES	LATENCY
01	027	D	E	N	01	00:16
02	015	B	E	N	01	00:16
03	013	C	C	Y	01	00:13
04	028	E	E	Y	01	00:07
05	005	D	A	N	01	00:15
06	004	A	A	Y	01	00:10
07	026	A	A	Y	01	00:43
08	006	A	A	Y	01	00:08
09	014	D	B	N	01	00:14
10	003	B	B	Y	01	00:17

COMMENT:THIS WAS A GOOD INTRODUCTION TO COMPUTER TESTING

001 FILES FOUND

Complete Score files are printed alphabetically by course at the RJE on a weekly schedule for use by students or faculty. Individual subjects and class lists are available on the printer or terminals at any time if needed for checking accuracy of the records or for monitoring student progress. A sample Student Scores file is shown in Figure 8.

Figure 8

SAMPLE STUDENT SCORES FILE

STUDENT: KOWAL , ART ID NUM: 99

TEST TRIAL	U2	U1	U3
01	07	08	08
02		11	
03		11	
04			
05			
06			
07			
08			
09			
10			

Anonymous student suggestions can be printed on demand to provide feedback to the faculty and staff.

Figure 9

SAMPLE ANONYMOUS STUDENT SUGGESTIONS

```
**********************************************************************
10/03/77                                        12:46 PM.
UNIT: 02
023,009,016,003,002,026,010,008,022,015,
ON TEST NUMBER 2, #3, THE QUESTION WAS ON THE COMBINATION OF 2 FOODS, AN
THEY WANTED TO KNOW WHICH HAD THE BEST COMBINATION OF TASTE AND TEXTURE.
I FEEL THIS QUESTION IS UNFAIR, BECAUSE ONE OF THE CHIOCES MAY BE SOME-
THING I'VE EATEN ALL MY LIFE AND ENJOYED TOGETHER.
**********************************************************************
10/07/77                                        12:59 PM.
UNIT: 01
005,020,016,017,024,014,013,026,025,023,
IN REGARDS TO QUESTION 5- THE CHOICES WERE NOT CLEAR. 4 1/2 TSP COULD ME
**********************************************************************
```

Two item analyses are provided to faculty. The partial (see Figure 10) includes data summarizing correctness of responses. At the end of each semester, a complete item analyses include latency response data, subjects' final choices, and the number of times each item was used for interactive testing as shown in Figure 11.

Figure 10

SAMPLE ITEM ANALYSIS REPORT

ITEM/ANALYSIS 10/31/77

COURSE: ED660 TEST: 2 UNIT: 00

ITEM	CORRECT RESPONSE	RESPONSE FREQUENCY A	B	C	D	E	F	TOTAL	CORRECT FIRST TRY
1	B	0	2	0	0	0	0	2	2
2	C	1	1	4	0	0	0	6	6
3	A	1	0	0	0	0	0	1	1
4	C	0	0	2	0	0	0	2	2
5	C	0	1	1	0	0	0	2	2

Figure 11

SAMPLE COMPLETE ITEM ANALYSIS REPORT

ITEM NUMBER: 3

		FIRST TRY RESPONSES					
CORRECT ANSWER IS: A	A	B	C	D	E	F	G
TOTAL NUMBER OF TIMES ITEM USED: 12	10	1	0	1	0	0	0

LATENCIES FOR ALL RESPONSES: (SEC):

LATENCIES(SEC):
15. 16. 18. 27. 29. 41. 42. 42. 45. 56. 58. 218.

MEDIAN: 42
Q1: 23 Q3: 51
LOW: 15 HIGH: 218
MIDSPREAD: 28

Daily procedures for proctors include setting up the sign-in book, reviewing instructions from the testing coordinator and programmers, and connecting terminals to the Burroughs B7700. Currently, three different models of Beehive cathode ray tube (CRT) terminals are used. The program permits use of DECwriter Model 36A printing terminals, but none are used at this time. Passwords are changed daily for security purposes and to discourage inquisitive students who may have observed the sign-in procedure during a recovery sequence while they were testing.

Each student is provided a Student User Manual at the beginning of the semester. However, some need assistance at their first session. They are assigned to a terminal after the sign-in at the proctor desk for each session. The terminal display provides cues to assist in the log-on process and to prompt students to read any news items concerning their course. A help session is available at the terminal to provide additional instructions for the testing procedures. Students then proceed with their test and respond to each item at the terminal. Proctors refer students to teaching assistants if asked any substantive questions. Some instructors maintain office hours in a consulting room in the Center. Others prefer to advise students in their own offices. This consulting is an important feature of the System.

A PROCTOR'S MANUAL and a SYSTEMS MANUAL provide staff with off-line documentation for the CAT System. Both the supervisor and utility programs provide cues and prompts to assist authorized users with managing the system and producing reports.

RESULTS

Equipment and space allocations indicate that each terminal will provide approximately two tests per hour at the beginning of each semester. The number increases by about fifty percent at the end of any semester. Subjects report that they become more relaxed at the terminals and with the system. During 1974, only two-thirds of the subjects reacted favorably to the interactive testing. During 1976, favorable reaction approached ninety percent. Most negative comments result from equipment failure and wording of test items. Time for hardcopy testing remained stable at about twenty-five minutes per test session. Waiting time for proctors to score tests varied because of the backlog created during peak periods. In addition, time was needed by the testing supervisor to enter test scores.

Subjects were assigned varying reinforcement one semester. However, the experimental treatment had to be terminated because subjects objected that their unequal treatment would cause unjust comparison at grading time.

Grades for final examinations in one course showed greater improvement than did grades in the same course taught during the preceding semester in a traditional lecture and hourly examination mode. Faculty reported fewer gripe sessions with students, feel they have more time for preparing instructional materials, consult with students, and conduct their own research when they use computer assisted testing.

REFERENCES

1. Marler, Charles D., INSTRUCTOR'S GUIDE TO ACCOMPANY PHILOSOPHY AND SCHOOLING, Allyn and Bacon, Boston, 1975

2. Martuza, Victor R., APPLYING NORM-REFERENCED AND CRITERION-REFERENCED MEASUREMENTS IN EDUCATION, Allyn and Bacon, Boston, 1977